

**CDMTCS  
Research  
Report  
Series**

**Von Neumann Normalisation of a  
Quantum Random Number  
Generator**

**A. A. Abbott and C. S. Calude**  
University of Auckland, NZ

CDMTCS-392  
October 2010

Centre for Discrete Mathematics and  
Theoretical Computer Science

# Von Neumann Normalisation of a Quantum Random Number Generator\*

Alastair A. Abbott<sup>†</sup>    Cristian S. Calude<sup>‡</sup>

Department of Computer Science

University of Auckland

Private Bag 92019, Auckland, New Zealand

`www.cs.auckland.ac.nz/~{aabb009,cristian}`

May 7, 2012

## Abstract

In this paper we study von Neumann un-biasing normalisation for ideal and real quantum random number generators, operating on finite strings or infinite bit sequences. In the ideal cases one can obtain the desired un-biasing. This relies critically on the independence of the source, a notion we rigorously define for our model. In real cases, affected by imperfections in measurement and hardware, one cannot achieve a true un-biasing, but, if the bias “drifts sufficiently slowly”, the result can be arbitrarily close to un-biasing. For infinite sequences, normalisation can both increase or *decrease* the (algorithmic) randomness of the generated sequences.

A successful application of von Neumann normalisation—in fact, any un-biasing transformation—does exactly what it promises, *un-biasing*, one (among infinitely many) symptoms of randomness; it will not produce “true” randomness.

## 1 Introduction

The outcome of some individual quantum-mechanical events cannot in principle be predicted, so they are thought of as ideal sources of random numbers. An incomplete list of quantum phenomena used for random number generation include nuclear decay radiation sources [29], the quantum mechanical noise in electronic circuits known as shot noise [30], photons travelling through a semi-transparent mirror [23, 27, 31, 34, 36] or photon arrival times [39, 32, 4]. Our methods are primarily developed to address these latter photon-based quantum random number generators (QRNGs), one of the most direct and popular ways to generate QRNs, but many of our mathematical results will be applicable to other QRNGs.

Due to imperfections in measurement and hardware, QRNGs are biased and operate

---

\*An extended abstract has appeared in A. A. Abbott, C. S. Calude. Von Neumann normalisation and symptoms of randomness: An application to sequences of quantum random bits, in C. S. Calude, J. Kari, I. Petre, G. Rozenberg (eds.). *Proc. 10th International Conference Unconventional Computation*, Lecture Notes Comput. Sci. 6714, Springer, Heidelberg, 2011, 40–51.

<sup>†</sup>AA was in part supported by the CDMTCS and FP7 Grant PIRSES-2010-269151-RANPHYS.

<sup>‡</sup>CC was in part supported by the CDMTCS, UoA R&SL grant and FP7 Grant PIRSES-2010-269151-RANPHYS.

non-independently in their generation of bits, two symptoms of non-randomness [10].<sup>1</sup> The first and simplest technique for reducing bias was invented by von Neumann [38]. It considers pairs of bits, and takes one of three actions: a) pairs of equal bits are discarded; b) the pair 01 becomes 0; c) the pair 10 becomes 1. Contrary to wide spread claims, the technique works for some sources of bits, but not for all. The source of constantly biased bits is effectively transformed into one in which the probabilities of 0 and 1 are equal: 50% for each. As we shall show, a stronger property is true: the un-biasing works not only for bits but for all reasonably long bit-strings. However, if the bias is not constant the procedure does not work. Furthermore, the von Neumann procedure cannot assure “true randomness” in its output. We briefly emphasise that bias is a property of the *source* of bits which only gives the expected frequency of 0’s and 1’s in the strings produced.

To understand the behaviour of QRNGs we need to study the un-biasing transformations on both (finite) strings and (infinite) sequences of bits produced by the source. In this paper we will focus on von Neumann normalisation<sup>2</sup> because it is very simple, easy to implement, and (along with the more efficient iterated version due to Peres [26] for which the results will also apply) is widely used by current proposals for QRNGs [23, 24, 18, 31]. Similar or stronger normalisation procedures have been studied under additional assumptions in, for example, Blum [7] (the source is a finite Markov chain) or Santha and Vazirani [28] (the source is semi-random) or Vadhan [37]; such additional hypotheses are satisfied by some physical sources, like a zener diode, but not necessarily by quantum sources. The widespread use of von Neumann normalisation, however, warrants a proper understanding of its operational quality when used on non-ideal sources regardless of the existence of other methods.

The main results of this paper are the following. In the “ideal case”, the von Neumann normalised output of an independent constantly biased QRNG is the probability space of the uniform distribution (un-biasing). This result is true for both for finite strings and for the infinite sequences produced by QRNGs (the QRNG runs indefinitely in the latter case).

It is important to note that independence in the mathematical sense of multiplicity of probabilities is a model intended to correspond to the physical notion of independence of outcomes [20]. In order to study the theoretical behaviour of QRNGs, which are based on the *assumption of physical independence of measurements*, we must translate this appropriately into our formal model. We carefully define independence of QRNGs to achieve this aim.

As explained above, QRNGs do not operate in ideal conditions. We develop a model for a real-world QRNG in which the bias, rather than holding steady, drifts slowly (within some bounds). In this framework we evaluate the speed of drift required to be maintained by the source distribution to guarantee that the output distribution is as close as one wishes to the uniform distribution.

We have also examined the effect von Neumann normalisation has on various properties of infinite sequences. In particular, Borel normality and (algorithmic) randomness are invariant under normalisation, but for  $\varepsilon$ -random sequences with  $0 < \varepsilon < 1$ , normalisation can both decrease or increase the randomness of the source.

Finally, we present our results in a mathematical framework which avoids hasty claims which later are disproved.

---

<sup>1</sup>As discussed in [1], “true randomness” does not mathematically exist. Various forms of algorithmic randomness [15] are each defined by an infinity of conditions, some “statistical” (like bias), some “non-statistical” (like lack of computable correlations).

<sup>2</sup>Many improvements of the scheme have been proposed [16, 26].

## 2 Notation

We present the main notation used throughout the paper.

By  $2^X$  we denote the power set of  $X$ . By  $|X|$  we denote the cardinality of the set of  $X$ .

Let  $B = \{0, 1\}$  and denote by  $B^*$  the set of all bit-strings ( $\lambda$  is the empty string). If  $x \in B^*$  and  $i \in B$  then  $|x|$  is the length of  $x$  and  $\#_i(x)$  represents the number of  $i$ 's in  $x$ . By  $B^n$  we denote the finite set  $\{x \in B^* \mid n = |x|\}$ . The concatenation product of two subsets  $X, Y$  of  $B^*$  is defined by  $XY = \{xy \mid x \in X, y \in Y\}$ . If  $X = \{x\}$  then we write  $xY$  instead of  $\{x\}Y$ . By  $B^\omega$  we denote the set of all infinite binary sequences. For  $\mathbf{x} \in B^\omega$  and natural  $n$  we denote by  $\mathbf{x}(n)$  the prefix of  $\mathbf{x}$  of length  $n$ . We write  $w \sqsubset v$  or  $w \sqsubset \mathbf{x}$  in case  $w$  is a prefix of the string  $v$  or the sequence  $\mathbf{x}$ .

A prefix-free (Turing) machine is a Turing machine whose domain is a prefix-free set of strings [10]. The prefix complexity of a string,  $H_W(\sigma)$ , induced by a prefix-free machine  $W$  is  $H_W(\sigma) = \min\{|p| : W(p) = \sigma\}$ . Fix a computable  $\varepsilon$  with  $0 < \varepsilon \leq 1$ . An  $\varepsilon$ -universal prefix-free machine  $U$  is a machine such that for every machine  $W$  there is a constant  $c$  (depending on  $U$  and  $W$ ) such that  $\varepsilon \cdot H_U(\sigma) \leq H_W(\sigma) + c$ , for all  $\sigma \in B^*$ . If  $\varepsilon = 1$  then  $U$  is simply called a universal prefix-free machine. A sequence  $\mathbf{x} \in B^\omega$  is called  $\varepsilon$ -random if there exists a constant  $c$  such that  $H_U(\mathbf{x}(n)) \geq \varepsilon \cdot n - c$ , for all  $n \geq 1$ . Sequences that are 1-random are simply called random.

A sequence  $\mathbf{x}$  is called Borel  $m$ -normal ( $m \geq 1$ ) if for every  $1 \leq i \leq 2^m$  one has:  $\lim_{n \rightarrow \infty} N_i^m(\mathbf{x}(n)) / \lfloor \frac{n}{m} \rfloor = 2^{-m}$ ; here  $N_i^m(y)$  counts the number of non-overlapping occurrences of the  $i$ th (in lexicographical order) binary string of length  $m$  in the string  $y$ . The sequence  $\mathbf{x}$  is called Borel normal if it is Borel  $m$ -normal, for every natural  $m \geq 1$ .

A probability space is a measure space such that the measure of the whole space is equal to one [6]. More precisely, a (Kolmogorov) probability space is a triple consisting of a sample space  $\Omega$ , a  $\sigma$ -algebra  $\mathcal{F}$  on  $\Omega$ , and a probability measure  $P$ , i.e. a countably additive function defined on  $\mathcal{F}$  with values in  $[0, 1]$  such that  $P(\Omega) = 1$ .

## 3 The finite case

### 3.1 Source probability space and independence

In this section we define the QRNG source probability space and the independence property.

Consider a string of  $n$  independent bits produced by a (biased) QRNG. Let  $p_0, p_1$  be the probability that a bit is 0 or 1, respectively, with  $p_0 + p_1 = 1$ ,  $p_0, p_1 \leq 1$ .

The probability space of bit-strings produced by the QRNG is  $(B^n, 2^{B^n}, P_n)$  where  $P_n : 2^{B^n} \rightarrow [0, 1]$  is defined by

$$P_n(X) = \sum_{x \in X} p_0^{\#_0(x)} p_1^{\#_1(x)}, \quad (1)$$

for all  $X \subseteq B^n$ .

It is easy to verify that the Kolmogorov axioms are satisfied for the space  $(B^n, 2^{B^n}, P_n)$ , so we have:

**Fact 1.** *The space  $(B^n, 2^{B^n}, P_n)$  with  $P_n$  defined in (1) is a probability space.*

The space  $(B^n, 2^{B^n}, P_n)$  is just the  $n$ -fold product of the single bit probability space  $(B, 2^B, P_1)$ . For this reason this space is often called an “independent identically-distributed bit source”. The resulting space is “independent” because each bit is independent of previous ones. But what is “an independent probability space”?

Physically the independence of a QRNG is usually expressed as the impossibility of extracting any information from the flow of bits  $x_1, \dots, x_{k-1}$  to improve chances of predicting the value of  $x_k$ , other than what one would have from knowing the probability space. The fact that photon-based QRNGs obey this physical independence between photons (and thus generated bits) rather well [2, 31] is the primary motivation for our modelling of these devices. These sources (where the condition of independence still holds) are often termed “independent-bit sources” [37]. In a real device we cannot, of course, expect each bit to be identically distributed, so we study this more general case more thoroughly in Section 3.5.

Formally, two events  $A, B \subseteq B^n$  are independent (in a probability space) if the probability of their intersection coincides with the product of their probabilities [9] (a complexity-theoretic approach was developed in [14]). This motivates the definition of independence of a general source probability space given in Definition 3. But first we need the following simple property:

**Fact 2.** *For every bit-string  $x$  and non-negative integers  $n, k$  such that  $0 \leq k + |x| \leq n$  we have:*

$$P_n(B^k x B^{n-k-|x|}) = p_0^{\#_0(x)} p_1^{\#_1(x)} = P_{|x|}(\{x\}). \quad (2)$$

**Definition 3.** The probability space  $(B^n, 2^{B^n}, \text{Prob}_n)$  is *independent* if for all  $1 \leq k \leq n$  and all  $x_1 \dots x_k \in B^k$  the events  $x_1 x_2 \dots x_{k-1} B^{n-k+1}$  and  $B^{k-1} x_k B^{n-k}$  are independent, i.e.

$$\text{Prob}_n(x_1 x_2 \dots x_{k-1} x_k B^{n-k}) = \text{Prob}_n(x_1 x_2 \dots x_{k-1} B^{n-k+1}) \cdot \text{Prob}_n(B^{k-1} x_k B^{n-k}).$$

**Fact 4.** *The probability space  $(B^n, 2^{B^n}, P_n)$  with  $P_n$  defined in (1) is independent.*

*Proof.* Using (2) we have:

$$\begin{aligned} P_n(x_1 x_2 \dots x_{k-1} x_k B^{n-k}) &= p_0^{\#_0(x_1 \dots x_k)} p_1^{\#_1(x_1 \dots x_k)} \\ &= p_0^{\#_0(x_1 \dots x_{k-1})} p_1^{\#_1(x_1 \dots x_{k-1})} p_0^{\#_0(x_k)} p_1^{\#_1(x_k)} \\ &= P_n(x_1 x_2 \dots x_{k-1} B^{n-k+1}) \cdot P_n(B^{k-1} x_k B^{n-k}). \end{aligned}$$

□

As we will see later, there are other relevant independent probability spaces.

### 3.2 Von Neumann normalisation function

Here we present formally the von Neumann normalisation procedure.

We define the mapping  $F : B^2 \rightarrow B \cup \{\lambda\}$  as

$$F(x_1 x_2) = \begin{cases} \lambda & \text{if } x_1 = x_2, \\ x_1 & \text{if } x_1 \neq x_2, \end{cases}$$

and  $f : B \rightarrow B^2$  as

$$f(x) = x\bar{x},$$

where  $\bar{x} = 1 - x$ . Note that for all  $x \in B$  we have  $F(f(x)) = x$  and, for all  $x_1, x_2 \in B$  with  $x_1 \neq x_2$ ,  $f(F(x_1x_2)) = x_1x_2$ .

For  $m \leq \lfloor n/2 \rfloor$  we define the normalisation function  $VN_{n,m} : B^n \rightarrow \left(\bigcup_{k \leq m} B^k\right) \cup \{\lambda\}$  as

$$VN_{n,m}(x_1 \dots x_n) = F(x_1x_2)F(x_3x_4) \cdots F\left(x_{(2\lfloor \frac{m}{2} \rfloor - 1)}x_{2\lfloor \frac{m}{2} \rfloor}\right).$$

**Fact 5.** For all  $1 < m \leq \lfloor n/2 \rfloor$  and  $y \in B^m$  there exists an  $x \in B^n$  such that  $y = VN_{n,m}(x)$ .

*Proof.* Take  $x = f(y_1)f(y_2) \cdots f(y_m)0^{n-2m}$ . □

In fact we can define the right inverse normalisation  $VN_{n,m}^{-1} : 2^{B^m} \rightarrow 2^{B^n}$  as

$$VN_{n,m}^{-1}(Y) = \left\{ u_1f(y_1)u_2f(y_2) \cdots u_mf(y_m)u_{m+1}v \mid y = y_1 \dots y_m \in Y, \right. \\ \left. u_i \in \{00, 11\}^*, v \in B \cup \{\lambda\}, |v| + 2m + \sum_{i=1}^{m+1} |u_i| = n \right\},$$

for which  $VN_{n,n}(VN_{n,m}^{-1}(y)) = \{y\}$  holds for every  $y \in B^m$ .

### 3.3 Target probability space and normalisation

We now construct the target probability space of the normalised bit-strings over  $B^m$  for  $m \leq \lfloor n/2 \rfloor$ , i.e. the probability space of the output bit-strings produced by the application of the von Neumann function on the output bit-strings generated by the QRNG.

The von Neumann normalisation function  $VN_{n,m}$  transforms the source probability space  $(B^n, 2^{B^n}, P_n)$  into the target probability space  $(B^m, 2^{B^m}, P_{n \rightarrow m})$ . The target space of normalised bit-strings of length  $1 < m \leq \lfloor n/2 \rfloor$  associated to the source probability space  $(B^n, 2^{B^n}, P_n)$  is the space  $(B^m, 2^{B^m}, P_{n \rightarrow m})$ , where  $P_{n \rightarrow m} : 2^{B^m} \rightarrow [0, 1]$  is defined for all  $Y \subseteq B^m$  by the formula:

$$P_{n \rightarrow m}(Y) = \frac{P_n(VN_{n,m}^{-1}(Y))}{P_n(VN_{n,m}^{-1}(B^m))}.$$

**Proposition 6.** The target space  $(B^m, 2^{B^m}, P_{n \rightarrow m})$  of normalised bit-strings of length  $1 < m \leq \lfloor n/2 \rfloor$  associated to the source probability space  $(B^n, 2^{B^n}, P_n)$  is a probability space.

*Proof.* We need to check only additivity: For  $X, Y \subseteq B^m$ ,  $X \cap Y = \emptyset \implies P_{n \rightarrow m}(X \cup Y) = P_{n \rightarrow m}(X) + P_{n \rightarrow m}(Y)$ . This equality is valid since  $VN_{n,m}^{-1}(X \cup Y) = VN_{n,m}^{-1}(X) \cup VN_{n,m}^{-1}(Y)$  and  $P_n(VN_{n,m}^{-1}(Y) \cup VN_{n,m}^{-1}(X)) = P_n(VN_{n,m}^{-1}(Y)) + P_n(VN_{n,m}^{-1}(X))$ , as  $VN_{n,m}^{-1}(X) \cap VN_{n,m}^{-1}(Y) = \emptyset$  because  $X$  and  $Y$  are disjoint. □

### 3.4 Normalisation of the output of a source with constant bias

We now show that von Neumann procedure transforms the source probability space with constant bias into the probability space with the uniform distribution over  $B^m$ , i.e. the target probability space  $(B^m, 2^{B^m}, P_{n \rightarrow m})$  has  $P_{n \rightarrow m} = U_m$ , the uniform distribution. Independence and the constant bias of  $P_n$  play a crucial role.

**Theorem 7** (von Neumann). *Assume that  $1 < m \leq \lfloor n/2 \rfloor$ . In the target probability space  $(B^m, 2^{B^m}, P_{n \rightarrow m})$  associated to the source probability space  $(B^n, 2^{B^n}, P_n)$  we have  $P_{n \rightarrow m}(Y) = U_m(Y) = |Y| \cdot 2^{-m}$ , for every  $Y \subseteq B^m$ .*

*Proof.* Since  $P_{n \rightarrow m}$  is additive it suffices to show that for any  $y \in B^m$ ,  $P_{n \rightarrow m}(\{y\}) = 2^{-m}$ . Let  $Z = P_n(VN_{n,m}^{-1}(B^m))$ .

We have (the sums are over all  $u_i \in \{00, 11\}^*$ ,  $v \in B \cup \{\lambda\}$  such that  $|v| + \sum_{i=1}^{m+1} |u_i| = n - 2m$ ):

$$\begin{aligned} P_{n \rightarrow m}(\{y\}) &= \frac{1}{Z} \sum_{u_i, v} p_0^{\#_0(u_1 f(y_1) \dots u_m f(y_m) u_{m+1} v)} p_1^{\#_1(u_1 f(y_1) \dots u_m f(y_m) u_{m+1} v)} \\ &= \frac{p_0^{\#_0(f(y_1) \dots f(y_m))} p_1^{\#_1(f(y_1) \dots f(y_m))}}{Z} \sum_{u_i, v} p_0^{\#_0(u_1 \dots u_{m+1} v)} p_1^{\#_1(u_1 \dots u_{m+1} v)} \\ &= \frac{p_0^m p_1^m}{Z} \sum_{u_i, v} p_0^{\#_0(u_1 \dots u_{m+1} v)} p_1^{\#_1(u_1 \dots u_{m+1} v)}, \end{aligned}$$

which is independent of  $y$ . Since  $P_{n \rightarrow m}(B^m) = 1$  and for all  $x_1, x_2 \in B^m$  we have  $P_{n \rightarrow m}(\{x_1\}) = P_{n \rightarrow m}(\{x_2\})$  it follows that  $P_{n \rightarrow m}(\{y\}) = 2^{-m} = U_m(\{y\})$ ; by additivity, for every  $Y \subseteq 2^m$  we have  $P_{n \rightarrow m}(Y) = U_m(Y) = |Y| \cdot 2^{-m}$ .  $\square$

It is natural to check whether the independence and constant bias of the source probability space are essential for the validity of the von Neumann normalisation procedure.

**Example 8.** The source probability space  $(B^2, 2^{B^2}, \text{Prob}_2)$  where  $\text{Prob}_2(00) = 0, \text{Prob}_2(01) = \text{Prob}_2(10) = \text{Prob}_2(11) = 1/3$  is independent and  $\text{Prob}_{2 \rightarrow 1} = U_1$ .

**Example 9.** The source probability space  $(B^2, 2^{B^2}, \text{Prob}_2)$  where  $\text{Prob}_2(00) = \text{Prob}_2(11) = 0, \text{Prob}_2(01) = 1/3, \text{Prob}_2(10) = 2/3$  is independent but  $\text{Prob}_{2 \rightarrow 1} \neq U_1$ .

**Comment.** One could present the above examples in the more general framework of Theorem 7.

**Theorem 10.** *Let  $m \geq 1$  and  $n = 2m$ . Consider the source probability space  $(B^n, 2^{B^n}, \text{Prob}_n) = \Pi_{i=1}^m(B^2, 2^{B^2}, P_2^i)$ , where  $P_2^i(01) = P_2^i(10)$ , for all  $1 \leq i \leq m$ . Then, in the target probability space  $(B^m, 2^{B^m}, \text{Prob}_{n \rightarrow m})$ , where  $\text{Prob}_n = \Pi_{i=1}^m P_2^i$ , we have  $\text{Prob}_{n \rightarrow m} = U_m$ .*

*Proof.* It is easy to check that for every  $y = y_1 \dots y_m \in B^m$  we have  $\text{Prob}_{n \rightarrow m}(\{y_1 \dots y_m\}) = \prod_{i=1}^m P_2^i(y_i \bar{y}_i) / \text{Prob}_n(VN_{n,m}^{-1}(B^m))$ , so  $\text{Prob}_{n \rightarrow m}(\{y_1 \dots y_m\})$  does not depend on  $y$  (because  $P_2^i(a\bar{a}) = P_2^i(\bar{a}a)$ , for every  $a \in B$ ). Hence,  $\text{Prob}_{n \rightarrow m} = U_m$ .  $\square$

The source probability space  $(B^n, 2^{B^n}, \text{Prob}_n)$  in Theorem 10 is not constantly biased and may be independent or not, but von Neumann normalisation still produces the uniform distribution under these conditions.

**Example 11.** The source probability space  $(B^4, 2^{B^4}, \text{Prob}_4)$  as in Theorem 10 where  $P_2^1(00) = P_2^1(01) = 1/3, P_2^1(10) = 1/4, P_2^1(11) = 1/12$  and  $P_2^2(00) = 1/12, P_2^2(01) = 1/4, P_2^2(10) = P_2^2(11) = 1/3$  is not independent and  $\text{Prob}_{4 \rightarrow 2} = U_2$ .

The outcome of successive context preparations and measurements, such as is the case for the type of QRNG usually envisioned, are postulated to be independent of previous and future outcomes [19]. This means there must be no causal link between one measurement and the next within the system (preparation and measurement devices included) so that the system has no memory of previous or future events. For QRNGs this translates into the condition that the probability that each successive bit is either 0 or 1 is independent of the previous bit measured. We will only consider such independent probability spaces, as this is a necessary property of a good RNG, so most QRNGs are designed to conform to this requirement.

The above assumption needs to be made clear as in high bit-rate experimental configurations to generate QRNs with, e.g., photons, its validity may not always be clear. If the wave-functions of successive photons “overlap” the assumption no longer holds and (anti)bunching phenomena may play a role. This is an issue that needs to be more seriously considered in QRNG design and will only become more relevant as the bit-rate of QRNGs is pushed higher and higher. While we leave study of the nature of these temporal correlations (and any non-independence they may cause) to future research [2], we pose the following open question which may help to quantify any possible effect they may have.

**Open Question.** Fix an integer  $k \geq 0$  and small positive real  $\kappa$ . Consider the probability space  $(B^n, 2^{B^n}, P_n^\dagger)$  where  $P_n^\dagger$  is a modification of the probability  $P_n$  satisfying the conditions that for all  $i \leq n$  and  $x_i \in B$  we have  $P_n(B^{i-1}x_iB^{n-i}) = P_n^\dagger(B^{i-1}x_iB^{n-i})$ ,

$$\left| P_n^\dagger(B^{i-1}x_iB^{n-i}) - P_n^\dagger(B^{i-1}x_iB^{n-i} \mid B^{i-k-1}x_{i-k} \dots x_{i-1}B^{n-i-1}) \right| \leq \kappa,$$

and for all  $l > k$

$$\begin{aligned} & P_n^\dagger(B^{i-1}x_iB^{n-i} \mid B^{i-l-1}x_{i-l} \dots x_{i-1}B^{n-i-1}) \\ &= P_n^\dagger(B^{i-1}x_iB^{n-i} \mid B^{i-k-1}x_{i-k} \dots x_{i-1}B^{n-i-1}). \end{aligned}$$

In other words, the probability of each bit depends on no more than the previous  $k$  bits, and the difference in probabilities for a bit between that given by  $P_n^\dagger$  conditioned on the previous  $k$  bits and  $P_n$  is no more than  $\kappa$ . *If the output of such a source is normalised with the von Neumann procedure, how close is the resulting probability space of strings of length  $m$  to the uniform distribution (see Definition 18 for a definition of the closeness of probability spaces)?*

### 3.5 Normalisation of the output of a source with non-constant bias

Now we consider the probability distribution obtained if von Neumann normalisation is applied to a string generated from an independent source with a non-constant bias—an “independent-bit source”. We consider only a bias which varies smoothly; this excludes the effects of sudden noise which could make the bias jump significantly from one bit to the next. Such a source corresponds to a QRNG in which the bias varies slowly (drifts)



from bit to bit over time, but never too far from its average point. We choose this to model photon-based QRNGs since the primary cause of variation in the bias will be of this nature. For example, the detector efficiencies may vary as a result of slow changes in temperature or power supply. While abrupt changes—which this model does not account for—are plausible, their relatively rare occurrence (in comparison with the bit generation rate in the order of MHz) will mean they have little effect on the resultant distribution.

Let  $p_0, p_1 < 1$  and  $p_0 + p_1 = 1$  be constant. Let  $x = x_1 x_2 \dots x_n \in B^n$  be the generated string. Then define the probability of an individual bit  $x_i$  being either zero or one as

$$q_i^{x_i} = \begin{cases} p_0 - \varepsilon_i & \text{if } x_i = 0, \\ p_1 + \varepsilon_i & \text{if } x_i = 1. \end{cases} \quad (3)$$

The variation in the bias is bounded, so we require that for all  $i$ ,

$$|\varepsilon_i| \leq \beta, \text{ with } \beta < \min(p_0, p_1).$$

Let  $\gamma_i = \varepsilon_{i+1} - \varepsilon_i$ . Furthermore, we assume that the “speed” of variation be bounded, i.e. there exists a positive  $\delta$  such that

$$|\gamma_i| \leq \delta, \quad (4)$$

for all  $i$ . Evidently we have  $\delta \leq \beta$  (presumably in any real situation  $\delta \ll \beta$ ); however, we introduce two separate constants since they correspond to two physically different (but related) concepts. Note that we will discuss in more detail the importance of these two parameters for the approximation of the uniform distribution and their relevance to calibration of the QRNG later once the analysis is completed. Indeed, the rate of change,  $\gamma_i$ , is more important; the need for  $\beta$  stems from the need to realise that, even though the probabilities can fluctuate, they can only fluctuate in one direction for so long (since  $q_i \in [0, 1]$ ), hence  $|\sum_i \gamma_i| = |\varepsilon_n - \varepsilon_1| \leq 2\beta$ .

For a string  $y = y_1 y_k \dots y_k \in B^k$  and positive integer  $i$  we introduce, for convenience, the following notation:

$$q_i(y) = q_i^{y_1} q_{i+1}^{y_2} \dots q_{i+k-1}^{y_k}.$$

The following fact will allow us to evaluate the effect of normalisation on such a string.

**Fact 12.** *The difference in probability between 01 and 10 depends only on  $\gamma_i$ , i.e.  $q_i(01) - q_i(10) = \gamma_i$ .*

*Proof.*

$$\begin{aligned} q_i(01) - q_i(10) &= (p_0 - \varepsilon_i)(p_1 + \varepsilon_{i+1}) - (p_1 + \varepsilon_i)(p_0 - \varepsilon_{i+1}) \\ &= (p_0 + p_1)(\varepsilon_{i+1} - \varepsilon_i) \\ &= \gamma_i. \end{aligned}$$

□

Let us first formally define the probability space generated by this QRNG.

**Proposition 13.** *The probability space of bit-strings produced by the QRNG is  $(B^n, 2^{B^n}, R_n)$  where  $R_n : 2^{B^n} \rightarrow [0, 1]$  is defined for all  $X \subseteq B^n$  as follows:*

$$R_n(X) = \sum_{x \in X} q_1(x). \quad (5)$$

*Proof.* We verify only that  $R_n(B^n) = 1$ , which is easily shown since  $q_i^0 + q_i^1 = 1$ , and  $R_n(B^n) = (q_1^0 + q_1^1) \cdots (q_n^0 + q_n^1)$ .  $\square$

**Fact 14.** For all  $i \geq 1$  and  $x, y \in \{0, 1\}^*$  we have:  $q_i(xy) = q_i(x)q_{i+|x|}(y)$ .

**Fact 15.** For all  $k, n \geq 1$ ,  $x \in \{0, 1\}^*$  with  $0 \leq k + |x| \leq n$  we have:

$$R_n(B^{n-k}xB^{n-k-|x|}) = q_{n-k+1}(x). \quad (6)$$

*Proof.* Using Fact 14 we get:

$$\begin{aligned} R_n(B^{n-k}xB^{n-k-|x|}) &= \sum_{y \in B^{n-k}} \sum_{z \in B^{n-k-|x|}} q_1(yxz) \\ &= \sum_{y \in B^{n-k}} \sum_{z \in B^{n-k-|x|}} q_1(y)q_{|y|+1}(x)q_{|y|+|x|+1}(z) \\ &= q_{n-k+1}(x) \sum_{y \in B^{n-k}} \sum_{z \in B^{n-k-|x|}} q_1(y)q_{|y|+|x|+1}(z) \\ &= q_{n-k+1}(x) \sum_{y \in B^{n-k}} q_1(y) \left( \sum_{z \in B^{n-k-|x|}} q_{|y|+|x|+1}(z) \right) \\ &= q_{n-k+1}(x). \end{aligned}$$

$\square$

**Fact 16.** The probability space  $(B^n, 2^{B^n}, R_n)$  with  $R_n$  defined in (5) is independent.

*Proof.* Using (6) we have:

$$\begin{aligned} R_n(x_1x_2 \dots x_{k-1}x_kB^{n-k}) &= q_1(x_1x_2 \dots x_{k-1}x_k) \\ &= q_1(x_1x_2 \dots x_{k-1})q_k(x_k) \\ &= R_n(x_1x_2 \dots x_{k-1}B^{n-k+1}) \cdot R_n(B^{k-1}x_kB^{n-k}). \end{aligned}$$

$\square$

As with the constantly biased source, we consider the probability space  $R_{n \rightarrow m}$ . We first investigate the simplest case  $n = 2m$ . In this situation, for any  $y \in B^m$  we have  $VN_{n,m}^{-1}(\{y\}) = \{f(y_1)f(y_2) \cdots f(y_m)\}$  and  $VN_{n,m}^{-1}(B^m) = \{f(z_1)f(z_2) \cdots f(z_m) \mid z = z_1 \dots z_m \in B^m\}$ .

**Fact 17.** The probability space of normalised bit-strings of length  $m = n/2$  is  $(B^m, 2^{B^m}, R_{n \rightarrow m})$  where  $R_{n \rightarrow m} : 2^{B^m} \rightarrow [0, 1]$  is defined for all  $Y \subseteq B^m$  as follows:

$$R_{n \rightarrow m}(Y) = \frac{R_n(VN_{n,m}^{-1}(Y))}{R_n(VN_{n,m}^{-1}(B^m))} = \sum_{y \in Y} \prod_{i=1}^m \frac{q_{2i-1}(f(y_i))}{q_{2i-1}(01) + q_{2i-1}(10)}. \quad (7)$$

### 3.6 Approximating the uniform distribution

Unlike the case for a constantly biased source, we no longer have  $q_i(01) = q_i(10)$ ; from Fact 12 we have  $q_i(01) = q_i(10) + \gamma_i$ . As a result the normalised equation is no longer the uniform distribution, but only an approximation thereof. We now explore how closely  $R_{n \rightarrow m}$  approximates  $U_m$ .

We first need to define what we mean by approximating  $U_m$ .

**Definition 18.** The *total variation distance* between two probability measures  $P$  and  $Q$  over the space  $\Omega$  is  $\Delta(P, Q) = \max_{A \subseteq \Omega} |P(A) - Q(A)|$ . We say that  $P$  and  $Q$  are  $\rho$ -close if  $\Delta(P, Q) \leq \rho$ .

It is well known (see for example [37]) that

**Lemma 19.** For finite  $\Omega$  we have  $\Delta(P, Q) = \frac{1}{2} \sum_{x \in \Omega} |P(\{x\}) - Q(\{x\})|$ .

The variation  $\Delta(R_{n \rightarrow m}, U_m)$  depends on each  $\gamma_i$  and  $q_i$  (thus on  $p_0, p_1$  and each  $\varepsilon_i$ ), but we wish to calculate the worst case in terms of the bounds  $\delta, \beta$  and  $p_0, p_1$ , i.e. using Lemma 19,

$$\max_{\gamma_i, q_i} \Delta(R_{n \rightarrow m}, U_m) = \frac{1}{2} \max_{\gamma_i, q_i} \sum_{y \in B^m} |R_{n \rightarrow m}(\{y\}) - 2^{-m}|.$$

Let us first note that we can write

$$\begin{aligned} \frac{q_{2i-1}(f(y_i))}{q_{2i-1}(01) + q_{2i-1}(10)} &= \frac{q_{2i-1}(f(y_i))}{2q_{2i-1}(f(y_i)) - (-1)^{y_i} \gamma_{2i-1}} \\ &= \frac{1}{2} \left( 1 + \frac{(-1)^{y_i} \gamma_{2i-1}}{2q_{2i-1}(f(y_i)) - (-1)^{y_i} \gamma_{2i-1}} \right), \end{aligned}$$

and hence we have

$$R_{n \rightarrow m}(\{y\}) = 2^{-m} \prod_{i=1}^m \left( 1 + \frac{(-1)^{y_i} \gamma_{2i-1}}{q_{2i-1}(01) + q_{2i-1}(10)} \right).$$

We have rewritten the denominator in its original form to emphasise that only the signs  $(-1)^{y_i}$  depend on  $y$ . Thus, we want to find the values of  $q_{2i-1}$  and  $\gamma_{2i-1}$  which maximise

$$\sum_{y \in B^m} \left| 1 - \prod_{i=1}^m \left( 1 + \frac{(-1)^{y_i} \gamma_{2i-1}}{q_{2i-1}(01) + q_{2i-1}(10)} \right) \right|, \quad (8)$$

subject to the constraints that  $|\gamma_\ell| \leq \delta$  and  $|\varepsilon_\ell| \leq \beta$  for  $1 \leq \ell \leq n$ .

**Lemma 20.** The function

$$g(c_1, \dots, c_n) = \sum_{y \in B^n} \left| \prod_{i=1}^n (1 + (-1)^{y_i} c_i) - 1 \right|$$

is strictly increasing for  $0 \leq c_i < 1$ ,  $i = 1, \dots, n$  (note that for  $1 \leq i \leq n$ ,  $g(c_1, \dots, c_i, \dots, c_n) = g(c_1, \dots, -c_i, \dots, c_n)$ ).

*Proof.* We take  $0 \leq c_i < 1$  for  $1 \leq i \leq n$ . For  $y = y_1 \dots y_n \in B^n$  define  $p(y, j) = \prod_{i=1, i \neq j}^n (1 + (-1)^{y_i} c_i)$ . Without loss of generality pick a  $j \leq n$  and let  $\varepsilon > 0$  be an (arbitrarily small) positive real with  $c_j + \varepsilon \leq 1$ . Note that

$$g(c_1, \dots, c_n) = \sum_{y \in B^n} |(1 + (-1)^{y_j} c_j) p(y, j) - 1|.$$

We partition  $B^n$  as follows:

$$\begin{aligned} Y_1 &= \{y \mid (1 - c_j - \varepsilon) p(y, j) - 1 \geq 0\}, \\ Y_2 &= \{y \mid (1 - c_j - \varepsilon) p(y, j) - 1 < 0 \text{ and } (1 - c_j) p(y, j) - 1 \geq 0\}, \\ Y_3 &= \{y \mid (1 - c_j) p(y, j) - 1 < 0 \text{ and } (1 + c_j) p(y, j) - 1 \geq 0\}, \\ Y_4 &= \{y \mid (1 + c_j) p(y, j) - 1 < 0 \text{ and } (1 + c_j + \varepsilon) p(y, j) - 1 \geq 0\}, \\ Y_5 &= \{y \mid (1 + c_j + \varepsilon) p(y, j) - 1 < 0\}. \end{aligned}$$

Note that for  $y \in B^n$ ,  $p(y, j) \geq 0$ , and for  $y_i \in Y_i$ ,  $i = 1, \dots, 5$ , we have

$$p(y_5, j) < p(y_4, j) < p(y_3, j) < p(y_2, j) < p(y_1, j),$$

and  $\bigcup_{i=1}^5 Y_i = B^n$ . We have:

$$\begin{aligned} g(c_1, \dots, c_j + \varepsilon, \dots, c_n) &= \sum_{i=1}^5 \sum_{y \in Y_i} |(1 + (-1)^{y_j} c_j + (-1)^{y_j} \varepsilon) p(y, j) - 1| \\ &= \sum_{y \in Y_1} [(1 + (-1)^{y_j} c_j) p(y, j) - 1 + (-1)^{y_j} \varepsilon p(y, j)] \\ &\quad + \sum_{i=2}^4 \sum_{y \in Y_i} (-1)^{y_j} [(1 + (-1)^{y_j} c_j) p(y, j) - 1 + (-1)^{y_j} \varepsilon p(y, j)] \\ &\quad + \sum_{y \in Y_5} -[(1 + (-1)^{y_j} c_j) p(y, j) - 1 + (-1)^{y_j} \varepsilon p(y, j)] \\ &= \sum_{i=1}^5 \sum_{y \in Y_i} |(1 + (-1)^{y_j} c_j) p(y, j) - 1| + 2\varepsilon \sum_{i=2}^4 \sum_{y \in Y_i} p(y, j) \\ &\quad - 2 \sum_{y \in Y_2} [(1 - c_j) p(y, j) - 1] + 2 \sum_{y \in Y_4} [(1 + c_j) p(y, j) - 1] \\ &= g(c_1, \dots, c_j, \dots, c_n) + 2\varepsilon \sum_{y \in Y_3} p(y, j) \\ &\quad - 2 \sum_{y \in Y_2} [(1 - c_j - \varepsilon) p(y, j) - 1] + 2 \sum_{y \in Y_4} [(1 + c_j + \varepsilon) p(y, j) - 1] \\ &> g(c_1, \dots, c_j, \dots, c_n), \end{aligned}$$

where the final line follows from the definition of  $Y_2$  and  $Y_4$ . Since this holds for all  $j \leq n$ ,  $g$  is strictly increasing over  $[0, 1]^n$ .  $\square$

Hence in order to maximise (8) we need to maximise the functions

$$u_j(\varepsilon_j, \gamma_j) = \left| \frac{\gamma_j}{q_j(01) + q_j(10)} \right| = \left| \frac{\gamma_j}{(p_0 - \varepsilon_j)(p_1 + \varepsilon_j + \gamma_j) + (p_1 + \varepsilon_j)(p_0 - \varepsilon_j - \gamma_j)} \right|, \quad (9)$$

for  $j = 2i - 1$ ,  $1 \leq i \leq m$ , subject to the constraints  $|\gamma_j| \leq \delta$ ,  $|\varepsilon_j| \leq \beta$  and  $|\varepsilon_{j+1}| = |\varepsilon_j + \gamma_j| \leq \beta$ .

**Lemma 21.** For every  $j \geq 1$  we have

$$u_j(\varepsilon_j, \gamma_j) \leq \begin{cases} u_j(\beta, -\delta) = u_j(\beta - \delta, \delta) & \text{if } p_1 \geq p_0, \\ u_j(-\beta, \delta) = u_j(-\beta + \delta, -\delta) & \text{if } p_0 > p_1, \end{cases} \quad (10)$$

$$= \frac{\delta}{2[p_0 p_1 - \beta(\beta - \delta) - |p_0 - p_1|(\beta - \delta/2)]}. \quad (11)$$

*Proof.* We omit the index  $j$  as it is not needed in this context. Let

$$v(\varepsilon, \gamma) = \frac{\gamma}{(p_0 - \varepsilon)(p_1 + \varepsilon + \gamma) + (p_1 + \varepsilon)(p_0 - \varepsilon - \gamma)}.$$

Since  $q(01) + q(10) > 0$ , in order to maximise  $u$  we look for maxima and minima of  $v$ ; clearly maxima have  $\gamma > 0$  and minima have  $\gamma < 0$ . We use Lagrange multipliers with inequality constraints to find the critical points. We have the following six constraints:  $h_1(\varepsilon, \gamma) = \varepsilon - \beta \leq 0$ ,  $h_2(\varepsilon, \gamma) = -\varepsilon - \beta \leq 0$ ,  $h_3(\varepsilon, \gamma) = \varepsilon + \gamma - \beta \leq 0$ ,  $h_4(\varepsilon, \gamma) = -\varepsilon - \gamma - \beta \leq 0$ ,  $h_5(\varepsilon, \gamma) = \gamma - \delta \leq 0$ ,  $h_6(\varepsilon, \gamma) = -\gamma - \delta \leq 0$ . We must solve the following equations:

$$\nabla_{\varepsilon, \gamma} v(\varepsilon, \gamma) + \sum_{i=1}^6 \lambda_i \nabla_{\varepsilon, \gamma} h_i(\varepsilon, \gamma) = 0, \quad (12)$$

$$\lambda_i h_i(\varepsilon, \gamma) = 0 \quad \text{for } i = 1, \dots, 6, \quad (13)$$

$$h_i(\varepsilon, \gamma) \leq 0 \quad \text{for } i = 1, \dots, 6, \quad (14)$$

$$\begin{cases} \lambda_i \geq 0 & \text{for minima, } i = 1, \dots, 6, \\ \lambda_i \leq 0 & \text{for maxima, } i = 1, \dots, 6. \end{cases} \quad (15)$$

We say a constraint is inactive if  $\lambda_i = 0$  and active otherwise; the condition of complementarity (13) captures the notion that a critical point satisfying the constraints either occurs at  $h_i(\varepsilon, \gamma) = 0$  or is also a critical point in the unconstrained problem.

Noting that  $0 < p_0 - \beta \leq p_0 + \beta < 1$  and solving, we find the candidate points are:

$$(\varepsilon, \gamma) = \begin{cases} (\frac{1}{2}(p_0 - p_1) \pm \frac{\delta}{2}, \mp \delta) \\ (\beta, -\delta), (\beta - \delta, \delta) & \text{for } p_0 - p_1 \leq 2\beta - \delta, \\ (-\beta, \delta), (-\beta + \delta, -\delta) & \text{for } p_1 - p_0 \leq 2\beta - \delta. \end{cases}$$

Note that  $u(\varepsilon, \gamma) = u(\varepsilon + \gamma, -\gamma)$ . Testing values shows the second case maximises  $u(\varepsilon, \gamma)$  when  $p_1 > p_0$  and the third cases maximises  $u(\varepsilon, \gamma)$  for  $p_0 > p_1$ . For  $p_0 = p_1$  both cases give the same value. Substituting in  $\varepsilon, \gamma$  and consolidating the cases we arrive at (11).  $\square$

Next we let

$$\alpha = \max_{\gamma_i, \varepsilon_i} u_j(\varepsilon_j, \gamma_j),$$

where  $u_j(\varepsilon_j, \gamma_j)$  comes from (9).

Then we have

$$\begin{aligned} \max_{\gamma_i, \varepsilon_i} \Delta(R_{n \rightarrow m}, U_m) &= \frac{1}{2} \sum_{y \in B^m} \left| \prod_{i=1}^m \left( \frac{1}{2} + (-1)^{y_i} \frac{\alpha}{2} \right) - 2^{-m} \right| \\ &= \frac{1}{2} \sum_{k=0}^m \binom{m}{k} \left| \left( \frac{1}{2} + \frac{\alpha}{2} \right)^k \left( \frac{1}{2} - \frac{\alpha}{2} \right)^{m-k} - 2^{-m} \right|. \end{aligned}$$

Note that in this worst case, the normalised source acts as an independent and identically-distributed source with  $p_0 = 1/2 \pm \alpha/2$  and the total variation is bounded by that of two binomial sources: one with  $p_0 = 1/2$ , the other with  $p_0 = 1/2 \pm \alpha/2$  (the number  $k$  of successful outcomes is identified with the number of ones in  $y$ ).

There are two interesting questions: a) what is the quality of the distribution produced by a QRNG, i.e. how close are  $R_{n \rightarrow m}$  and  $U_m$  in terms of  $\alpha$ ? and b) given a real  $\rho \in (0, 1)$ , how accurate does the QRNG need to be in terms of  $\alpha$  to guarantee that  $R_{n \rightarrow m}$  and  $U_m$  are  $\rho$  close?

We can take a rough approach to solve the above problems as follows. First note that

$$\begin{aligned} \Delta(R_{n \rightarrow m}, U_m) &\leq \frac{1}{2} \sum_{y \in B^m} \left| \prod_{i=1}^m \left( \frac{1}{2} + (-1)^{y_i} \frac{\alpha}{2} \right) - 2^{-m} \right| \\ &\leq \frac{1}{2} \sum_{y \in B^m} \frac{1}{2^m} ((1 + \alpha)^m - 1) \\ &= \frac{1}{2} ((1 + \alpha)^m - 1). \end{aligned}$$

So given  $\alpha$ ,  $R_{n \rightarrow m}$  and  $U_m$  are at most  $\frac{1}{2} ((1 + \alpha)^m - 1)$ -close. Conversely,  $R_{n \rightarrow m}$  and  $U_m$  are  $\rho$  close if

$$\alpha \leq (1 + 2\rho)^{1/m} - 1. \quad (16)$$

We will express further results in the latter form, focusing on question b), although both are important questions depending on the operational circumstances and results can easily be transformed from one form to the other.

So, by making  $\alpha$  very small,  $R_{n \rightarrow m}$  can be made as close as we wish to the uniform distribution. This is intuitive since  $\alpha \rightarrow 0$  only as  $\delta \rightarrow 0$  and we approach the constantly biased source situation.

There are, unfortunately, some issues with this bound. First, as  $m \rightarrow \infty$  the bound on the variation becomes infinite too. This is unreasonable as by definition we should have  $\Delta(R_{n \rightarrow m}, U_m) \leq 1$ . It only makes sense to talk about  $\rho \leq 1$ , although in any useful situation we will require  $\rho$  to be small (close to 0) so it is only of real importance that the bound is good in this situation. However, (16) requires  $\alpha$  to be significantly smaller than we really require for the two probabilities to be  $\rho$  close. Even for small  $\rho$  the bound is no-way near tight enough (see Figure 2). Further, it would be instructive to examine more correctly the behaviour for large  $m$  and investigate fully the nature of the relationship between  $\alpha$ ,  $m$  and  $\rho$ .

To rectify this and find a more reasonable bound, we carry out a finer analysis making use of the previous observation that this is the same problem as finding the variation between two binomial distributions. Let us denote a binomial probability distribution function for  $n$  trials and probability of success  $p$  as  $S_{n,p} : \{0, \dots, n\} \rightarrow [0, 1]$  where for each  $A \subseteq \{0, \dots, n\}$ ,

$$S_{n,p}(A) = \sum_{k \in A} \binom{n}{k} p^k (1-p)^{n-k}.$$

For  $0 \leq p, p' \leq 1$ , we then have

$$\Delta(S_{n,p}, S_{n,p'}) = \frac{1}{2} \sum_{k=0}^n \binom{n}{k} \left| p^k (1-p)^{n-k} - (p')^k (1-p')^{n-k} \right|,$$

and

$$\max_{\gamma_i, \varepsilon_i} \Delta(R_{n \rightarrow m}, U_m) = \Delta(S_{m, 1/2(1 \pm \alpha)}, S_{m, 1/2}).$$

**Fact 22.** For  $0 \leq p, p' \leq 1$  we have  $\Delta(S_{n,p}, S_{n,p'}) = \Delta(S_{n,1-p}, S_{n,1-p'})$ .

The total variation between two binomial distributions can be given in terms of regularised incomplete beta functions [3].

**Definition 23.** The *incomplete beta function* is defined as

$$B_\ell(a, b) = \int_0^\ell u^{a-1} (1-u)^{b-1} du.$$

For  $\ell = 1$  we write  $B_1(a, b) = B(a, b)$  for the *complete beta function*, or just *beta function*. The *regularised incomplete beta function* is defined as

$$I_\ell(a, b) = \frac{B_\ell(a, b)}{B(a, b)}.$$

**Theorem 24.** Let  $0 \leq p \leq 1$ ,  $q = 1 - p$  and  $0 \leq x \leq q$ . The total variation between two binomial distributions with probability of success  $p$  and  $p + x$  is

$$\begin{aligned} \Delta(S_{n,p}, S_{n,p+x}) &= n \int_p^{p+x} S_{n-1,u}(\ell-1) du \\ &= n \binom{n-1}{\ell-1} \int_p^{p+x} u^{\ell-1} (1-u)^{n-\ell} du \\ &= I_{p+x}(\ell, n-\ell+1) - I_p(\ell, n-\ell+1), \end{aligned}$$

where

$$\lceil np \rceil \leq \ell := \ell(n, p, x) = \left\lceil \frac{-n \log(1-x/q)}{\log(1+x/p) - \log(1-x/q)} \right\rceil \leq \lceil n(p+x) \rceil.$$

*Proof.* The first line is from Adell and Jodrá [3]. The rest follows from the well known properties of the beta functions:  $B_\ell(a, b) = B_\ell(b, a)$  and

$$\binom{n}{k} = \frac{1}{(n+1)B(n-k+1, k+1)}.$$

□

**Theorem 25.** The total variation is bounded by

$$\begin{aligned} \Delta(R_{n \rightarrow m}, U_m) &\leq \Delta(S_{m, 1/2}, S_{m, 1/2(1+\alpha)}) \\ &= I_{1/2(1+\alpha)}(\ell, m-\ell+1) - I_{1/2}(\ell, m-\ell+1), \\ &= F(m-\ell; m, 1/2-\alpha/2) - F(m-\ell; m, 1/2) \end{aligned}$$

where

$$\lceil m/2 \rceil \leq \ell = \ell(m, 1/2, \alpha/2) = \left\lceil \frac{-m \log(1-\alpha)}{\log(1+\alpha) - \log(1-\alpha)} \right\rceil \leq \lceil m(1+\alpha)/2 \rceil,$$

and

$$F(k; n, p) = \sum_{x=0}^k S_{n,p}(x)$$

is the cumulative distribution function for the binomial distribution.

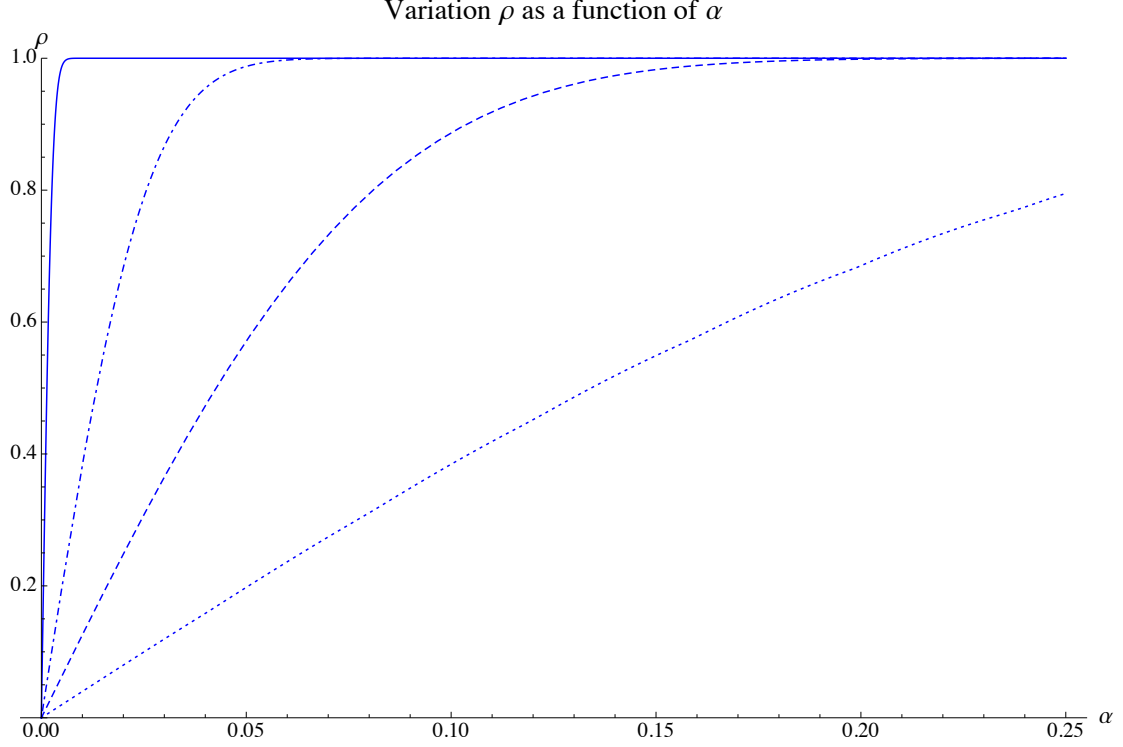


Figure 1: Plot of  $\rho$  against  $\alpha$  using the bound in Theorem 25 for four values of  $m$ : 100 (dotted), 1,000 (dashed), 10,000 (dot-dashed) and 1,000,000 (solid).

*Proof.* This follows directly from Theorem 24 and Fact 22. The last line follows from well known properties of the binomial distribution.  $\square$

This bound is exact (under the extrema given by Lemma 21), and we easily verify that  $\Delta(R_{n \rightarrow m}, U_m) \leq 1$  since  $I_p(a, b) \leq 1$  for all  $a, b$  and  $p \leq 1$ , and for  $p' \geq p$  we have  $I_{p'}(a, b) \geq I_p(a, b)$  (with equality only for  $p = p'$ ). Unfortunately this bound on the variation has no simple closed form, so we can not easily relate  $\alpha$ ,  $m$  and  $\rho$  like we did in (16). The shape and nature of this relationship can be seen for various values of  $m$  in Figure 1. In practice, with  $m$  fixed and given  $\rho$  it is easy to compute (with numerical methods)  $\alpha$  such that  $\Delta(R_{n \rightarrow m}, U_m) \leq \rho$ . For relatively small  $\rho$  however, we can find a simple and fairly good bound which is easy to work with for rough approximations.

**Theorem 26.** Assume that  $m = n/2$ . Consider the probability spaces  $(B^m, 2^{B^m}, R_{n \rightarrow m})$  and  $(B^m, 2^{B^m}, U_m)$ . For every real  $\rho$  such that  $0 \leq \rho < 1$ , if

$$\alpha \leq \rho \sqrt{\frac{2\pi(1 - \frac{2}{m})}{m+1}},$$

then  $\Delta(R_{n \rightarrow m}, U_m) \leq \rho$ .

*Proof.* We will take a first order (linear) approximation of  $\Delta(S_{m,1/2}, S_{m,1/2(1+\alpha)})$  around  $\alpha = 0$ . From Theorem 24 and the Fundamental Theorem of Calculus we have

$$\Phi(\alpha) := \frac{d}{d\alpha} \Delta(S_{m,1/2}, S_{m,1/2(1+\alpha)}) = m \binom{m-1}{\ell-1} 2^{-m} (1+\alpha)^{\ell-1} (1-\alpha)^{m-\ell}.$$

Since  $\ell \geq \lceil m/2 \rceil$  we have

$$\Phi(\alpha) \leq \Phi(0),$$



so our first order upper bound is given by

$$\Delta(S_{m,1/2}, S_{m,1/2(1+\alpha)}) \leq \alpha \Phi(0) = \alpha m \binom{m-1}{\ell-1} 2^{-m}.$$

Since the central binomial coefficient (i.e.  $\binom{n}{\lfloor n/2 \rfloor}$ ) is the largest, for  $k \leq m-1$  we have

$$\binom{m-1}{k} \leq \binom{m-1}{\lfloor \frac{m-1}{2} \rfloor} = \binom{m-1}{\lceil \frac{m}{2} \rceil - 1},$$

which can easily be shown by taking the two cases of  $m$  odd and  $m$  even. Since  $\ell \geq \lceil m/2 \rceil$  we have that

$$\Phi(0) \leq 2^{-m} m \binom{m-1}{\lceil \frac{m}{2} \rceil - 1} = 2^{-m} m \frac{\lceil \frac{m}{2} \rceil}{m} \binom{m}{\lceil \frac{m}{2} \rceil} = 2^{-m} \lceil m/2 \rceil \binom{m}{\lceil \frac{m}{2} \rceil}.$$

Using the bounds given in Corollary 2.3, [33], and writing  $m = a \lceil m/2 \rceil$  where  $a \leq 2$ , we have

$$\begin{aligned} \binom{a \lceil \frac{m}{2} \rceil}{\lceil \frac{m}{2} \rceil} &< \frac{1}{\sqrt{2\pi \lceil \frac{m}{2} \rceil}} \frac{a^{m+\frac{1}{2}}}{(a-1)^{(a-1)\lceil \frac{m}{2} \rceil + \frac{1}{2}}} \\ &= \frac{1}{\sqrt{2\pi \lceil \frac{m}{2} \rceil}} \frac{m^{m+\frac{1}{2}}}{\lfloor \frac{m}{2} \rfloor^{\lfloor \frac{m}{2} \rfloor + \frac{1}{2}} \lceil \frac{m}{2} \rceil^{\lceil \frac{m}{2} \rceil}} \\ &\leq \frac{1}{\sqrt{2\pi \lceil \frac{m}{2} \rceil}} \frac{m^{m+\frac{1}{2}}}{((\frac{m}{2} + \frac{1}{2})(\frac{m}{2} - \frac{1}{2}))^{\lfloor \frac{m}{2} \rfloor} (\frac{m}{2} - \frac{1}{2})^{\frac{1}{2}} (\frac{m}{2})} \\ &\leq \frac{1}{\sqrt{2\pi \lceil \frac{m}{2} \rceil}} \frac{2^{m+\frac{1}{2}}}{(1 - \frac{1}{m^2})^{\lfloor \frac{m}{2} \rfloor} (1 - \frac{1}{m})^{\frac{1}{2}}} \\ &\leq \frac{1}{\sqrt{\pi \lceil \frac{m}{2} \rceil}} \frac{2^m}{(1 - \frac{1}{2m})(1 - \frac{1}{m})^{\frac{1}{2}}} \\ &\leq \frac{2^m}{\sqrt{\pi \lceil \frac{m}{2} \rceil} (1 - \frac{2}{m})}. \end{aligned}$$

Hence, we have

$$\Phi(0) \leq \sqrt{\frac{\lceil \frac{m}{2} \rceil}{\pi(1 - \frac{2}{m})}} \leq \sqrt{\frac{m+1}{2\pi(1 - \frac{2}{m})}}.$$

□

This bound is much better than the bound given in (16), and for small  $\alpha$  is extremely good. It has the desired properties that as  $\alpha \rightarrow 0$ , the bound on the variation tends to 0 also. Obviously this bound is not less than one for all  $\alpha$ , but for small  $\rho$  the bound is very good, as can be seen in Figure 2.

Another interesting question refers to the possibility of manipulating the parameter  $\alpha$  for fine calibration of the QRNG. For  $R_{n \rightarrow m}$  to become closer to  $U_m$  we need to make  $\alpha$  smaller, but this can be done by adjusting both  $\delta$  and  $\beta$ . As previously discussed, both are reasonable physical parameters, and which one is the most suitable (or easiest) to

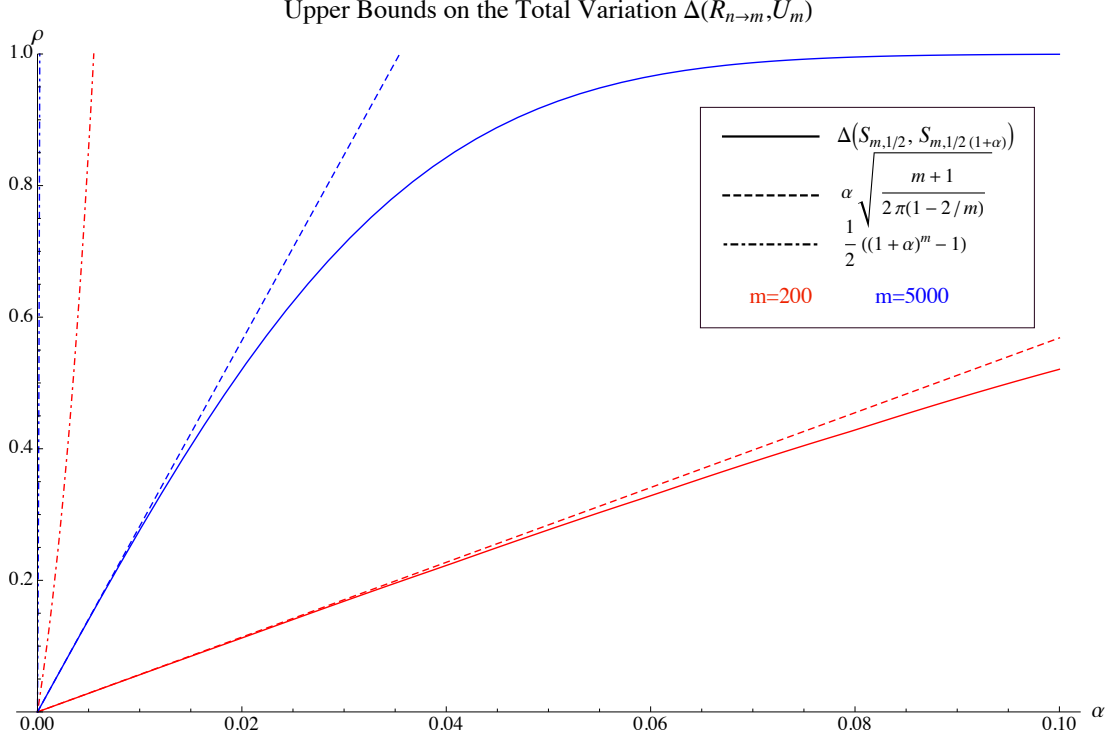


Figure 2: Plot of upper bounds on the variation between  $R_{n \rightarrow m}$  and  $U_m$ .

decrease experimentally will to a large extent depend on the QRNG set-up itself. However, adjusting  $\delta$  has a larger effect on  $\alpha$  than adjusting  $\beta$  does, and  $R_{n \rightarrow m}$  will only approach  $U_m$  arbitrarily close as  $\delta \rightarrow 0$ , as even with  $\beta = \delta$  (recall  $\delta \leq \beta$ ) we do not have  $\alpha = 0$  unless  $\delta = 0$ .

These results can be extended to all  $m \leq n/2$ , although the analysis is rather elaborated. The key difference is that in the definition of  $R_{n \rightarrow m}$  in (7) the set  $VN_{n,m}^{-1}(Y)$  no longer has the same size as  $Y$ , so an additional summation is needed in the right hand side of (7). However, the total variation will still be maximised under the same conditions as in Lemmata 20 and 21, and the same relation as in Theorem 25 holds.

It is worth noting that the conditions which maximised the variation in (10) correspond to every  $\varepsilon_i$  being the same up to a small variation  $\delta$ . Physically this would indicate that  $p_0, p_1$  have been incorrectly stated, but that the device is actually rather accurate except for a small drift in probabilities of no more than  $\delta$ . Since the parameters  $\varepsilon_i$  are supposed to physically account for the amount the probability is allowed to drift, which will normally be much more than the drift between individual bits (the  $\gamma_i$ ), if the device is calibrated so that  $p_0$  and  $p_1$  are centred so that the  $\varepsilon_i$  are distributed around them, then the variation will not be nearly as bad as in this worst case. However, the bound on the variation remains valid as it is not necessarily meaningful (or useful) to look into the physical situation under which the worst case bound is achieved.

We briefly wish to point out that other methods for dealing with independent-bit sources have been proposed. For example, grouping bits into blocks of size  $\ell$  and taking the parity of these bits for the “normalised” bit, produces a string of length  $n/\ell$  [37]. With this method each bit becomes unbiased exponentially fast in  $\ell$ . However, the bound in Theorem 26 is asymptotically tighter than the corresponding bound that can be obtained by the parity method if the block size  $\ell$  is fixed; if  $\ell$  scales polynomially with  $n$  then

this method produces a better bound, but at a substantial cost to the number of bits produced [37, Proposition 6.5]. The reason the von Neumann normalisation outperforms the parity method is due to the fact that the bias is required to vary slowly.

## 4 The infinite case

The extension of the above results to infinite sequences of bits produced by QRNGs is fairly straightforward, but forces us to address a few unexpected problems. First, we must extend the definition of the normalisation function  $VN_{n,m}$  to sequences. We define  $VN : B^\omega \rightarrow B^\omega \cup B^*$  as

$$VN(\mathbf{x} = x_1 \dots x_n \dots) = F(x_1 x_2) F(x_3 x_4) \cdots F(x_{2\lfloor \frac{n}{2} \rfloor - 1} x_{2\lfloor \frac{n}{2} \rfloor}) \cdots$$

For convenience we also define  $VN_n : B^\omega \rightarrow \left(\bigcup_{k \leq n} B^k\right) \cup \{\lambda\}$  as

$$VN_n(\mathbf{x}) = F(x_1 x_2) F(x_3 x_4) \cdots F(x_{2\lfloor \frac{n}{2} \rfloor - 1} x_{2\lfloor \frac{n}{2} \rfloor}) = VN_{n,n}(x_1 \dots x_n).$$

Secondly, we introduce the probability space of infinite sequences as in [10]. Let  $A_Q = \{a_1, \dots, a_Q\}$ ,  $Q \geq 2$  be an alphabet with  $Q$  elements. We let  $\mathcal{P} = \{xA_Q^\omega \mid x \in A_Q^*\} \cup \{\emptyset\}$  and  $\mathcal{C}$  be the class of all finite mutually disjoint unions of sets in  $\mathcal{P}$ ; the class  $\mathcal{P}$  can be readily shown to generate a  $\sigma$ -algebra  $\mathcal{M}$ . Using Theorem 1.7 from [10], the probabilities on  $\mathcal{M}$  are characterised by the functions  $h : A_Q^* \rightarrow [0, 1]$  satisfying:

1.  $h(\lambda) = 1$ ,
2.  $h(x) = h(x_{a_1}) + \cdots + h(x_{a_Q})$ , for all  $x \in A_Q^*$ .

If  $Q = 2$  so  $A_2 = B$ , and for  $x \in B^n$  we take  $h(x) = P_n(\{x\})$  with  $P_n$  as defined in Fact 1, then the above conditions are satisfied. This induces our probability measure  $\mu_P$  on  $\mathcal{M}$ , which satisfies  $\mu_P(XB^\omega) = P_n(X)$  for  $X \subseteq B^n$ . Hence the suitable extension of the finite case probability space to infinite generated sequences is the space  $(B^\omega, \mathcal{M}, \mu_P)$ . In the special case when  $p_0 = p_1$  we get the Lebesgue probability  $\mu_{PL}(XB^\omega) = \sum_{x \in X} 2^{-|x|}$ .

In general, if  $Q \geq 2$ ,  $p_i \geq 0$  for  $i = 1, \dots, Q$  are reals in  $[0, 1]$  such that  $\sum_{i=1}^Q p_i = 1$ , we can take  $h_Q(x) = p_1^{\#_{a_1}(x)} \cdots p_Q^{\#_{a_Q}(x)}$  ( $\#_{a_i}(x)$  is the number of occurrences of  $a_i$  in  $x$ ) to obtain the probability space  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$  in which  $\mu_{P_Q}(xA_Q^\omega) = h_Q(x)$ , for all  $x \in A_Q^*$ .

The first result notes that there exist sequences  $\mathbf{x} \in B^\omega$  such that  $VN(\mathbf{x}) \in B^*$ . In fact every string can be produced via von Neumann normalisation from a suitable sequence.

**Theorem 27.** *For every string  $y \in B^*$  there exists an uncountable set  $R \subset B^\omega$  of  $\mu_P$  measure zero such that for all  $\mathbf{x} \in R$ ,  $VN(\mathbf{x}) = y$ .*

*Proof.* Let  $y = y_1 \dots y_n \in B^*$  and  $D = \{00, 11\}$ , the two-bit blocks which are deleted by von Neumann normalisation and  $y' = f(y_1) \dots f(y_n)$ . Then every sequence  $\mathbf{x} \in y'D^\omega$  satisfies  $VN(\mathbf{x}) = VN_{2n}(\mathbf{x})VN(x_{2n+1}x_{2n+2} \dots) = y$  since  $VN_{2n}(\mathbf{x}) = VN_{2n, 2n}(y') = y$  and for all  $\mathbf{z} \in D^\omega$  we have  $VN(\mathbf{z}) = \lambda$ . Obviously, the set  $R = y'D^\omega$  is uncountable and has  $\mu_P$  measure zero as the set of Borel normal sequences has measure one [10].  $\square$

**Corollary 28.** *The set  $Q = \{\mathbf{x} \in B^\omega \mid VN(\mathbf{x}) \in B^*\}$  has  $\mu_P$  measure zero.*

*Proof.* We simply note that the union of countably many measure zero sets also has measure zero.  $\square$

It is interesting to note that the “collapse” in the generated sequence produced by von Neumann normalisation in Theorem 27 is not due to computability properties of the sequence. In particular, there are random sequences that collapse to any string, so to strings which are not Borel normal.

In the following we need a measure-theoretic characterisation of random sequences, so we present a few facts from constructive topology and probability.

Consider the compact topological space  $(A_Q^\omega, \tau)$  in which the basic open sets are the sets  $wA_Q^\omega$ , with  $w \in A_Q^*$ . Accordingly, an open set  $G \subset A_Q^\omega$  is of the form  $G = VA_Q^\omega$ , where  $V \subset A_Q^*$ .

From now on we assume that the reals  $p_i, 1 \leq i \leq Q$  which define the probability  $\mu_{P_Q}$  are all computable. A constructively open set  $G \subset A_Q^\omega$  is an open set  $G = VA_Q^\omega$  for which  $V \subset A_Q^*$  is computably enumerable (c.e.). A constructive sequence of constructively open sets, c.s.c.o. sets for short, is a sequence  $(G_m)_{m \geq 1}$  of constructively open sets  $G_m = V_m A_Q^\omega$  such that there exists a c.e. set  $X \subset A_Q^* \times \mathbb{N}$  with  $V_m = \{x \in A_Q^* \mid (x, m) \in X\}$ , for all natural  $m \geq 1$ . A constructively null set  $S \subset A_Q^\omega$  is a set for which there exists a c.s.c.o. sets  $(G_m)_{m \geq 1}$  with  $S \subset \bigcap_{m \geq 1} G_m$ ,  $\mu_{P_Q}(G_m) \leq 2^{-m}$ . A sequence  $\mathbf{x} \in A_Q^\omega$  is random in the probability space  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$  if  $\mathbf{x}$  is not contained in any constructively null set in  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$ . For the case of the Lebesgue probability  $\mu_{P_L}$  the measure-theoretic characterisation of random sequences holds true:  $\mathbf{x}$  is random if and only if  $\mathbf{x}$  is not contained in any constructively null set of  $(A_Q^\omega, \mathcal{M}, \mu_{P_L})$  [25, 10].

We continue with another instance in which von Neumann normalisation decreases randomness.

**Proposition 29.** *There exist (continuously many) infinite 1/2-random sequences  $\mathbf{x} \in B^\omega$  such that  $VN(\mathbf{x}) = 000\dots 00\dots$ .*

*Proof.* Consider a random sequence  $\mathbf{x} = x_1x_2\dots x_n\dots$  and construct the sequence  $\mathbf{x}' = 0x_10x_2\dots 0x_n\dots$ . Clearly,  $\mathbf{x}'$  is 1/2-random, but  $VN(\mathbf{x}') = 000\dots 00\dots$  because there exist infinitely many 1's in  $\mathbf{x}$ .  $\square$

We follow this with instances for which the converse is true: von Neumann normalisation conserves or increases randomness.

**Proposition 30.** *There exist (continuously many) infinite 1/2-random sequences  $\mathbf{x} \in B^\omega$  such that  $VN(\mathbf{x})$  is random.*

*Proof.* Consider a random sequence  $\mathbf{x} = x_1x_2\dots x_n\dots$  and construct the sequence  $\mathbf{x}' = x_1\bar{x}_1x_2\bar{x}_2\dots x_n\bar{x}_n\dots$ . Clearly,  $\mathbf{x}'$  is 1/2-random and  $VN(\mathbf{x}') = \mathbf{x}$ .  $\square$

**Comment.** Both Proposition 29 and 30 are true for the more general case of  $\varepsilon$ -random sequences, where  $0 < \varepsilon < 1$  is computable.

We briefly note that in the definition of Borel normality it does not matter if we count the number of non-overlapping occurrences of each string of length  $m$ ,  $N_i^m(y)$  as defined in Section 2, or the number of overlapping occurrences,  $\mathcal{N}_i^m(y)$  [22]. One of the main results of this section, presented in Theorem 31, is the following: Borel normality is invariant under von Neumann normalisation.

**Theorem 31.** *Let  $\mathbf{x} \in B^\omega$  be Borel normal in  $(B^\omega, \mathcal{M}, \mu_{P_L})$ . Then  $VN(\mathbf{x})$  is also Borel normal in  $(B^\omega, \mathcal{M}, \mu_{P_L})$ .*

*Proof.* Note that  $VN(\mathbf{x}) \in B^\omega$  because  $\mathbf{x}$  contains infinitely many occurrences of 01 on even/odd positions. Let  $D = \{00, 11\}$ ,  $\mathbf{x}^*(n) = VN_{n,n}(\mathbf{x}(n))$ ,  $n' = |\mathbf{x}^*(n)|$ . We have

$$\lim_{n' \rightarrow \infty} \frac{N_i^m(\mathbf{x}^*(n))}{n'} = \lim_{n' \rightarrow \infty} \left( \frac{n}{n'} \right) \left( \frac{N_i^m(\mathbf{x}^*(n))}{n} \right),$$

but as  $n \rightarrow \infty$ ,  $n' \rightarrow \infty$ . We thus have

$$\begin{aligned} \lim_{n' \rightarrow \infty} \frac{n'}{n} &= \lim_{n' \rightarrow \infty} \frac{N_0^1(\mathbf{x}^*(n)) + N_1^1(\mathbf{x}^*(n))}{n} \\ &= \lim_{n \rightarrow \infty} \frac{\mathcal{N}_{01}^2(\mathbf{x}(n)) + \mathcal{N}_{10}^2(\mathbf{x}(n))}{\lfloor n/2 \rfloor} \\ &= 2^{-1} \end{aligned}$$

by the normality of  $\mathbf{x}$ . The number of occurrences of each  $i = i_1 \dots i_m \in B^m$  in  $\mathbf{x}^*(n)$  is the number of occurrences of  $i' = f(i_1)y_1f(i_2) \dots y_{m-1}f(i_m)$  in  $\mathbf{x}(n)$ , summed over all  $y_1, \dots, y_{m-1} \in D^*$ . Viewing  $i'$  as a string over  $\{00, 01, 10, 11\}$  we have:

$$\begin{aligned} \lim_{n' \rightarrow \infty} \frac{N_i^m(\mathbf{x}^*(n))}{n} &= \lim_{n \rightarrow \infty} \frac{\sum_{y_1, \dots, y_{m-1}} N_{i'}^{|i'|}(\mathbf{x}(n))}{n} \\ &= \sum_{y_1 \in D^*} \sum_{y_2 \in D^*} \dots \sum_{y_{m-1} \in D^*} 2^{-2|i'|} \\ &= \sum_{|y_1|=0}^{\infty} 2^{|y_1|} \sum_{|y_2|=0}^{\infty} 2^{|y_2|} \dots \sum_{|y_{m-1}|=0}^{\infty} 2^{|y_{m-1}|} 2^{-2|i'|} \\ &= 2^{-2m} \sum_{|y_1|=0}^{\infty} 2^{-|y_1|} \sum_{|y_2|=0}^{\infty} 2^{-|y_2|} \dots \sum_{|y_{m-1}|=0}^{\infty} 2^{-|y_{m-1}|} \\ &= 2^{-2m} 2^{m-1} \\ &= 2^{-(m+1)}. \end{aligned}$$

Hence, both limits exist and we have

$$\begin{aligned} \lim_{n' \rightarrow \infty} \frac{N_i^m(\mathbf{x}^*(n))}{n'} &= \lim_{n' \rightarrow \infty} \left( \frac{n}{n'} \right) \left( \frac{N_i^m(\mathbf{x}^*(n))}{n} \right) \\ &= \frac{\lim_{n' \rightarrow \infty} \frac{N_i^m(\mathbf{x}^*(n))}{n}}{\lim_{n' \rightarrow \infty} \frac{n'}{n}} \\ &= \frac{2^{-(m+1)}}{2^{-1}} \\ &= 2^{-m}. \end{aligned}$$

Since this holds for all  $m, i$  we have that  $VN(\mathbf{x})$  is Borel normal.  $\square$

Let  $A_Q = \{a_1, \dots, a_Q\}$ ,  $Q \geq 3$ . Let  $\sum_{i=1}^Q p_i = 1$  where  $p_i \geq 0$  for  $i = 1, \dots, Q$  and  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$  be the probability space defined by the probabilities  $p_i$ . Let  $A_{Q-1} = \{a_1, \dots, a_{Q-1}\}$  and  $(A_{Q-1}^\omega, \mathcal{M}, \mu_{P_{Q-1}^T})$  be the probability space defined by the probabilities

$$p_i^T = p_i \left( 1 + \frac{p_Q}{\sum_{j=1}^{Q-1} p_j} \right) = \frac{p_i}{1 - p_Q},$$

with  $1 \leq i \leq Q-1$ . Let  $T : A_Q^* \rightarrow A_{Q-1}^*$  be the monoid morphism defined by  $T(a_i) = a_i$  for  $1 \leq i \leq Q-1$ ,  $T(a_Q) = \lambda$ ;  $T(x) = T(x_1)T(x_2) \cdots T(x_n)$  for  $x \in A_Q^n$ . As  $T$  is prefix-increasing we naturally extend  $T$  to sequences to obtain the function  $T : A_Q^\omega \rightarrow A_{Q-1}^\omega$  given by  $T(\mathbf{x}) = \lim_{n \rightarrow \infty} T(\mathbf{x}(n))$  for  $\mathbf{x} \in A_Q^\omega$ .

**Lemma 32.** *The transformation  $T$  is  $(\mu_{P_Q}, \mu_{P_{Q-1}^T})$ -preserving, i.e. for all  $w \in A_{Q-1}^*$  we have  $\mu_{P_Q}(T^{-1}(wA_{Q-1}^\omega)) = \mu_{P_{Q-1}^T}(wA_{Q-1}^\omega)$ .*

*Proof.* Take  $w = w_1 \dots w_m \in A_{Q-1}^\omega$ . We have:

$$\begin{aligned} \mu_{P_Q}(T^{-1}(wA_{Q-1}^\omega)) &= \mu_{P_Q}(\{\mathbf{x} \in A_Q^\omega \mid w \sqsubset T(\mathbf{x})\}) \\ &= \mu_{P_Q}\left\{a_Q^{i_1}w_1a_Q^{i_2}w_2 \dots a_Q^{i_m}w_m\mathbf{z} \mid \mathbf{z} \in A_Q^\omega\right\} \\ &= \sum_{i_1, \dots, i_m=0}^{\infty} h_Q\left(a_Q^{i_1}w_1a_Q^{i_2}w_2 \dots a_Q^{i_m}w_m\right) \\ &= \sum_{i_1, \dots, i_m=0}^{\infty} h_{Q-1}(w) \cdot p_Q^{i_1 + \dots + i_m} \\ &= h_{Q-1}(w) \cdot \frac{1}{1 - p_Q} \\ &= h_{Q-1}^T(w) \\ &= \mu_{P_{Q-1}^T}(wA_{Q-1}^\omega). \end{aligned}$$

□

**Proposition 33.** *If  $\mathbf{x} \in A_Q^\omega$  is random in  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$  and  $T$  is the transformation defined in Lemma 32, then  $T(\mathbf{x})$  is random in  $(A_{Q-1}^\omega, \mathcal{M}, \mu_{P_{Q-1}^T})$ .*

*Proof.* We generalise a result in [12] stating that, for the Lebesgue probability, measure-preserving transformations preserve randomness. Assume that  $\mathbf{x}$  is random in  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$  but  $T(\mathbf{x})$  is not random in  $(A_{Q-1}^\omega, \mathcal{M}, \mu_{P_{Q-1}^T})$ , i.e. there is a constructive null set  $R = (G_m)_{m \geq 1}$  containing  $T(\mathbf{x})$ . Assume that  $G_m = X_m A_{Q-1}^\omega$ , where  $X_m \subset A_{Q-1}^\omega$  is c.e. and has the measure  $\mu_{P_{Q-1}^T}(X_m A_{Q-1}^\omega)$  smaller than  $2^{-m}$ . Define  $S_m = T^{-1}(X_m A_{Q-1}^\omega) \subset A_Q^\omega$  and note that  $S_m$  is open because it is equal to  $\bigcup_{w \in X_m} V_w A_Q^\omega$  with  $V_w = \{v \in A_Q^\omega \mid w \sqsubset T(v)\}$  and, using Lemma 32, has the measure smaller than  $2^{-m}$ :

$$\begin{aligned} \mu_{P_Q}(S_m) &= \mu_{P_Q}\left(\bigcup_{w \in X_m} V_w A_Q^\omega\right) \\ &\leq \sum_{w \in X_m} \mu_{P_Q}(V_w A_Q^\omega) \\ &= \sum_{w \in X_m} \mu_{P_Q}(T^{-1}(wA_{Q-1}^\omega)) \\ &= \mu_{P_{Q-1}^T}(X_m A_{Q-1}^\omega) \\ &\leq 2^{-m}. \end{aligned}$$

We have proved that  $\mathbf{x}$  is not random in  $(A_Q^\omega, \mathcal{M}, \mu_{P_Q})$ , a contradiction. □

Let us define  $VN^{-1} : 2^{B^*} \rightarrow 2^{B^*}$  for  $x = x_1 \dots x_m \in B^m$  as

$$\begin{aligned} VN^{-1}(x) &= \{y \mid y = u_1 f(x_1) u_2 \dots u_m f(x_m) u_{m+1} v \text{ and} \\ &\quad u_i \in \{00, 11\}^* \text{ for } 1 \leq i \leq m, v \in B \cup \{\lambda\}\} \\ &= \bigcup_{n=0}^{\infty} VN_{n+2m, m}^{-1}(x), \end{aligned}$$

and for  $X \subseteq B^*$  as

$$VN^{-1}(X) = \bigcup_{x \in X} VN^{-1}(x).$$

For all  $x \in B^*$  and  $\mathbf{y} \in VN^{-1}(x)B^\omega$  we then have  $x \sqsubset VN(\mathbf{y})$ .

For the cases that  $VN(\mathbf{x}) \in B^\omega$ , the probability space  $(B^\omega, \mathcal{M}, \mu_{P_{VN}})$  induced by von Neumann normalisation is endowed with the measure  $\mu_{P_{VN}}$ . The measure  $\mu_{P_{VN}}$  is defined on the sets  $xB^\omega$  with  $x \in B^*$  by

$$\mu_{P_{VN}}(xB^\omega) = \frac{\mu_P(VN^{-1}(x)B^\omega)}{\mu_P(VN^{-1}(B^{|x|})B^\omega)}.$$

By noting that  $VN^{-1}(B^{|x|}) \subset VN^{-1}(B^*)$  it is clear to see that  $\mu_{P_{VN}}$  satisfies the Kolmogorov axioms for a probability measure. While the set  $VN^{-1}(B^{|x|})$  contains sequences for which normalisation produces a finite string, from Corollary 28 we know that the set of such sequences have measure zero, so the definition of  $\mu_{P_{VN}}$  is a good model of the target probability space. We thus arrive at the key result that (measure-theoretical) randomness is invariant under von Neumann normalisation.

**Theorem 34.** *Let  $\mathbf{x} \in B^\omega$  be random in  $(B^\omega, \mathcal{M}, \mu_P)$ . Then  $VN(\mathbf{x}) \in B^\omega$  is also random in  $(B^\omega, \mathcal{M}, \mu_{P_{VN}})$ .*

*Proof.* We write the random sequence  $\mathbf{x}$  as  $\mathbf{x} = x_1 x_2 \dots x_n \dots = (x_1 x_2) \dots (x_{2n-1} x_{2n}) \dots \in \{00, 01, 10, 11\}^\omega$ . Renaming  $a = 00, A = 01, B = 10, b = 11$  and consistently deleting first all occurrences of  $a$  we get a random sequence  $\mathbf{x}_{A,B,b}$  on the alphabet  $\{A, B, b\}$ , then deleting all occurrences of  $b$  we get a random sequence  $\mathbf{x}_{A,B}$  on the alphabet  $\{A, B\}$ . The result follows from the fact that  $VN(\mathbf{x}) = \mathbf{x}_{0,1}$  and Proposition 33 stating that  $\mathbf{x}_{A,B}$  is random.  $\square$

**Corollary 35.** *If  $\mathbf{x} \in B^\omega$  is random in  $(B^\omega, \mathcal{M}, \mu_P)$  then  $VN(\mathbf{x})$  is Borel normal in  $(B^\omega, \mathcal{M}, \mu_{P_{VN}})$ .*

*Proof.* From Theorem 34 it follows that  $VN(\mathbf{x})$  is Borel normal provided  $\mathbf{x}$  is random [10].  $\square$

**Theorem 36.** *The probability space  $(B^\omega, \mathcal{M}, \mu_{P_{VN}})$  induced by von Neumann normalisation is the uniform distribution  $(B^\omega, \mathcal{M}, \mu_{P_L})$ , where  $\mu_{P_L}$  is the Lebesgue measure.*

*Proof.* By Lemma 32 von Neumann normalisation is measure preserving, so for  $x \in B^*$  we have

$$\begin{aligned} \mu_{P_{VN}}(xB^\omega) &= \mu_P(VN^{-1}(x)B^\omega) \\ &= p_0^{|x|} p_1^{|x|} \sum_{d_i \in D^*} p_0^{\#_0(d_1 \dots d_{|x|})} p_1^{\#_1(d_1 \dots d_{|x|})}. \end{aligned}$$

The key point, as in the finite case, is that this only depends on  $|x|$  not  $x$  itself. By using the fact that for any  $n$ ,  $\sum_{x \in B^n} \mu_{P_{VN}}(xB^\omega) = 1$ , we have

$$\mu_{P_{VN}}(xB^\omega) = 2^{-|x|}$$

for all  $x \in B^*$ , and hence  $\mu_{P_{VN}} = \mu_{P_L}$ , the Lebesgue measure.  $\square$

This can easily be extended from the case when  $VN(\mathbf{x})$  is infinite, to the case in which it is finite. To do so, note that if  $\mathbf{y} \in B^\omega$  and  $VN(\mathbf{x}) = y \in B^n$ , then the probability space induced by von Neumann normalisation is  $(B^n, 2^{B^n}, P_n^*)$ . We then have

$$P_n^*(x) = \frac{\mu_P(VN^{-1}(x)D^\omega)}{\mu_P(VN^{-1}(B^n)D^\omega)},$$

and since the denominator is constant for all  $x \in B^n$ , we can proceed as for above, and  $P_n^* = U_n$  as desired.

**Theorem 37.** *The set  $\{\mathbf{x} \in B^\omega \mid VN(\mathbf{x}) \in B^* \text{ or } VN(\mathbf{x}) \in B^\omega \text{ is computable}\}$  has measure zero with respect to the probability space  $(B^\omega, \mathcal{M}, \mu_P)$ .*

*Proof.* By Theorem 34 we deduce that

$$\{\mathbf{x} \in B^\omega \mid VN(\mathbf{x}) \in B^\omega \text{ is computable}\} \subset \{\mathbf{x} \in B^\omega \mid \mathbf{x} \text{ is not random in } (B^\omega, \mathcal{M}, \mu_P)\},$$

which has measure zero [25]. To complete the proof, note that we know from Corollary 28 that the set  $\{\mathbf{x} \in B^\omega \mid VN(\mathbf{x}) \in B^*\}$  also has measure zero.  $\square$

## 5 Role of probability spaces for QRNGs

The treatment of QRNGs as entirely probabilistic devices is grounded purely on the probabilistic treatment of measurement in quantum mechanics which originated with Born's decision to "give up determinism in the world of atoms" [8], a viewpoint which has become a core part of our understanding of quantum mechanics. This is formalised by the Born rule, but the probabilistic nature of *individual* measurement is nonetheless postulated and tells us nothing about *how* the probability arises. Along with the assumption of independence this allows us to predict the probability of *successive* events, as we have done.

No-go theorems such as the Kochen-Specker Theorem [21] tell us something stronger: if we assume non-contextuality (i.e. that the result of an observation is independent of the compatible observables are co-measured alongside it [5, 17]) then there can, in general, be no pre-existing definite values prescribable to certain sets of measurement outcomes in dimension three or greater Hilbert space. In other words, the randomness is not due to ignorance of the system being measured; indeed, since there are in general no definite values associated with the measured observable it is surprising there is an outcome at all [35]. While this does not answer the question as to where the randomness arises from, it does tell us something stronger than the Born Rule does. In [13] it is shown that every infinite sequence produced by a QRNG is (strongly) uncomputable. In particular, this implies that it is *impossible* for a QRNG to output a computable sequence. The set of computable numbers has measure zero with respect the probability space of the QRNG, but the impossibility of producing such sequence is much stronger than, although not in contradiction with, the probabilistic results.



In the finite case every string is, of course, obtainable, and we would expect the distribution to be that predicted by the probability space derived from the Born Rule. However, the infinite case has something to say here too. We can view any finite string produced by a QRNG as the initial segment of an infinite sequence the QRNG would produce if left to run indefinitely. For any infinite sequence produced by the QRNG, it is impossible to compute the value of *any bit* before it is measured [1]; in the finite case this means there is no way to provably compute the value of the next bit before it is measured. In light of value indefiniteness this is not unexpected, but nonetheless gives mathematical grounding to the postulated unpredictability of each individual measurement, as well as the independence of successive measurements—indeed we can rule out any computable causal link within the system which may give rise to the measurement outcome.

The results we have presented in this paper, however, describe thoroughly the distribution of strings/sequences produced by QRNGs. With the distributions known we can create more intelligent tests of the quality of output of a QRNG [11]. Current statistical tests for analysing RNGs are designed with pseudo-RNGs in mind, and are not necessarily the best way to test the quality of QRNGs. The effects of normalisation on strings generated by QRNGs can help us design QRNGs which are more robust to experimental imperfection and exhibit the desired behaviour. It will further aid in developing new normalisation techniques designed to produce the expected (ideal) theoretical distribution even in the absence of experimental imperfections.

## 6 Conclusions

The analysis developed in this paper involves the probability spaces of the source and output of a QRNG and the effect von Neumann normalisation has on these spaces.

In the “ideal case”, the von Neumann normalised output of an independent constantly biased QRNG is the probability space of the uniform distribution (un-biasing). This result is true for both for finite strings and for the infinite sequences produced by QRNGs (the QRNG runs indefinitely in the second case).

For a real-world QRNG in which the bias, rather than holding steady, drifts slowly, we evaluated the speed of drift required to be maintained by the source distribution to guarantee that the output distribution is arbitrarily close to the uniform distribution. It is an *open question* to study the quality of von Neumann normalisation in the more realistic case when, instead of the bits being independent, the probability for each bit depends on a finite number of preceding bits (for example, because of the high bit-rate of the experiment). Note that Blum’s algorithm [7] assumes a Markov-type correlation, which cannot be assumed for a QRNG certified by value indefiniteness [2].

We have also examined the effect von Neumann normalisation has on various properties of infinite sequences. In particular, Borel normality and (algorithmic) randomness are invariant under normalisation, but for  $\varepsilon$ -random sequences with  $0 < \varepsilon < 1$ , normalisation can both decrease or increase the randomness of the source. It is an *open question* whether von Neumann normalisation preserves randomness and Borel normality for finite strings.

Finally, we reiterate that a successful application of von Neumann normalisation—in, fact, any un-biasing transformation—does exactly what it promises, *un-biasing*, one (among infinitely many) symptoms of randomness; it will not produce “true” randomness.

## Acknowledgment

We thank Karl Svozil and Marius Zimand for many discussions and suggestions on the topics of the paper, as well as the anonymous referees for suggestions which improved the presentation of the paper.

## References

- [1] A. A. Abbott, C. S. Calude, J. Conder, and K. Svozil. Unpublished work on the incomputability of quantum randomness, in preparation.
- [2] A. A. Abbott, C. S. Calude, and K. Svozil. A quantum random number generator certified by value indefiniteness. *CDMTCS Research Report*, 396, 2010; to appear in *Mathematical Structures in Computer Science*.
- [3] J. Adell and P. Jodrá. Exact Kolmogorov and total variation distances between some familiar discrete distributions. *Journal of Inequalities and Applications*, (64307), 2006.
- [4] H.-Q. M. and Y. Xie and L.-A. Wu. Random number generation based on the time of arrival of single photons. *Applied Optics*, 44(36):7760–7763, 2005.
- [5] J. Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, 38(3):447–452, 1966.
- [6] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, Toronto, London, 1979.
- [7] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite state Markov chain. *Combinatorica*, 6(2):97–108, 1986.
- [8] M. Born. Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 38:803–837, 1926. English translation by J. A. Wheeler and W. H. Zurek, in *Quantum Theory and Measurement*, chapter I.2. Princeton University Press, 1983.
- [9] P. Busch, P. J. Lahti, and P. Mittelstaedt. *The Quantum Theory of Measurement*. Lecture Notes in Physics: New Series m, Monographs; 2. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 1996.
- [10] C. S. Calude. *Information and Randomness: An Algorithmic Perspective*. Springer-Verlag, Berlin, 2nd edition, 2002.
- [11] C. S. Calude, M. J. Dinneen, M. Dumitrescu, and K. Svozil. Experimental evidence of quantum randomness incomputability. *Physical Review A*, 82(022102):1–8, 2010.
- [12] C. S. Calude, P. Hertling, H. Jürgensen, and K. Weihrauch. Randomness on full shift spaces. *Chaos, Solutions & Fractals*, 12(3):491–503, 2001.
- [13] C. S. Calude and K. Svozil. Quantum randomness and value indefiniteness. *Advanced Science Letters*, 1(165–168), 2008.
- [14] C. S. Calude and M. Zimand. Algorithmically independent sequences. *Information and Computation*, 208:292–308, 2010.
- [15] R. Downey and D. Hirschfeldt. *Algorithmic Randomness and Complexity*. Theory and Applications of Computability. Springer, 2010.

- [16] P. Elias. The efficient construction of an unbiased random sequence. *The Annals of Mathematical Statistics*, 43(3):865–870, 1972.
- [17] P. Heywood and M. L. G. Redhead. Nonlocality and the Kochen-Specker paradox. *Foundations of Physics*, 13(5):481–499, 1983.
- [18] id Quantique. Quantis - quantum random number generators. <http://idquantique.com/products/quantis.htm>, 12/08/2009.
- [19] J. M. Jauch. *Foundations of Quantum Mechanics*. Addison-Wesley, Reading, MA, 1968.
- [20] M. Kac. *Statistical Independence in Probability, Analysis and Number Theory*. The Carus Mathematical Monographs. The Mathematical Association of America, 1959.
- [21] S. Kochen and E. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87, 1967. Reprinted in E. Specker. *Selecta*. Birkhäuser Verlag, Basel, 1990.
- [22] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. John Wiley & Sons, New York, 1974.
- [23] O. Kwon, Y. Cho, and Y. Kim. Quantum random number generator using photon-number path entanglement. *Applied Optics*, 48(9):1774–1778, 2009.
- [24] H. Ma, S. Wang, D. Zhang, J. Change, L. Ji, Y. Hou, and L. Wu. A random-number generator based on quantum entangled photon pairs. *Chinese Physics Letters*, 21(19):1961–1964, 2004.
- [25] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.
- [26] Y. Peres. Iterating von Neumann’s procedure for extracting random bits. *The Annals of Statistics*, 20(1):590–597, 1992.
- [27] S. Pironia, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmchenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by Bell’s theorem. *Nature*, 464(09008), 2010.
- [28] M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.
- [29] H. Schmidt. Quantum-mechanical random-number generator. *Journal of Applied Physics*, 41(2):462–468, 1970.
- [30] Y. Shen, L. Tian, and H. Zou. Practical quantum random nubmer generator based on measuring the shot noise of vacuum states. *Physical Review A*, 81(063814), 2010.
- [31] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden. Optical quantum random number generator. *Journal of Modern Optics*, 47(4):595–598, 2000.
- [32] M. Stipčević and B. M. Rogina. Quantum random number generator based on photonic emission in semiconductors. *Review of Scientific Instruments*, 78(4):045104, 2007.

- [33] P. Stănică. Good lower and upper bounds on binomial coefficients. *Journal of Inequalities in Pure and Applied Mathematics*, 2(3):30, 2001.
- [34] K. Svozil. The quantum coin toss – testing microphysical undecidability. *Physics Letters A*, 143(9):433–437, 1990.
- [35] K. Svozil. Quantum information via state partitions and the context translation principle. *Journal of Modern Optics*, 51:811–819, 2004.
- [36] K. Svozil. Three criteria for quantum random-number generators based on beam splitters. *Physical Review A*, 79(5):054306, 2009.
- [37] S. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. now publishers, 2011. To appear. Current version at <http://people.seas.harvard.edu/~salil/pseudorandomness>, accessed on 5.5.2012.
- [38] J. von Neumann. Various techniques used in connection with random digits. *National Bureau of Standards Applied Math Series*, **12** (1951), 36–38. In A. H. Traub, editor, *John von Neumann, Collected Works*, pages 768–770. MacMillan, New York, 1963.
- [39] M. A. Wayne, E. R. Jeffrey, G. M. Aksolrod, and P. G. Kwiat. Photon arrival time quantum random number generation. *Journal of Modern Optics*, 56(4):516–522, 2009.